Am. J. Hum. Genet. 51:1–6, 1992

Invited Editorial:[1] The Human Genome Project: Where Did It Come From, Where Is It Going?

Elke Jordan

National Center for Human Genome Research, National Institutes of Health, Bethesda

Introduction

The Human Genome Project is an international research initiative with the goal of producing detailed genetic and physical maps of each of the 24 different human chromosomes and, when technology allows it to be done at a reasonable cost, determining the sequence of the 3 billion nucleotides that make up human DNA. The time frame needed to accomplish all this has been estimated to be 15 years, and the cost has been estimated to be $200 million per year, or $3 billion total. In the United States, the Human Genome Project is managed principally by two government agencies, the National Institutes of Health (NIH) and the Department of Energy (DOE).

In view of its mission to improve the health of all Americans, the NIH is naturally interested in this initiative as a foundation for future research in human genetics and biology. The DOE interest stems from a long-standing program of genetic research directed at improving the ability to assess the effects of radiation and energy-related chemicals on human health. In recognition of these related interests, as well as the complementary scientific strengths of the two agencies, NIH and DOE have agreed to coordinate their respective genome activities according to a Memorandum of Understanding that outlines plans for cooperation on genome research. This document was signed by both agencies on October 1, 1988.

Received February 24, 1992.

Address for correspondence and reprints: Elke Jordan, National Center for Human Genome Research, National Institutes of Health, Building 38A, Room 605, Bethesda, MD 20892.

This article represents the opinion of the author and has not been peer reviewed.

1. This editorial was prepared at the invitation of the ASHG Human Genome Committee.

Origins of the Human Genome Project

The possibility of mounting a Human Genome Project was discussed in the scientific community over a number of years and at a variety of meetings, dating back to at least 1985. A good overview of this period is provided by Watson (1990). Human geneticists had started serious attempts to isolate human disease genes, and it was clear that the task was enormously difficult because of the lack of good genetic or physical maps, such as those molecular biologists were accustomed to using in studies of simpler organisms. It became obvious that detailed reference maps of all human chromosomes and, ultimately, of the sequence of all human DNA would be a boon to searches for human genes and would allow scientists to focus on the profound biological questions that can be answered only when the gene is at hand. The prospect of having comprehensive information about the human genome was both intellectually exciting and recognized as holding enormous potential for progress in human biology and medicine. Because of the labor involved, it made sense to do the mapping and sequencing once and for all and in the most efficient and cost-effective fashion possible.

As discussions proceeded, ideas and strategies for a Human Genome Project matured and became more practical. This process was particularly aided by the work of two committees that developed highly influential reports. The first was the National Research Council (NRC) Committee on Mapping and Sequencing the Human Genome (National Research Council 1988) and the other was a committee organized by the U.S. Congress Office of Technology Assessment (OTA) (Office of Technology Assessment 1988). Both committees supported the concept of a Human Genome Project and recommended that the U.S. government provide funding for it. As a result of these deliberations, the current three-step approach to ac-

complishing the goals of the Human Genome Project became accepted. In brief, this approach is to begin with genetic and physical mapping of chromosomes and to proceed to sequencing of the DNA when improved and cheaper methods are available. Technology development to improve all the techniques that were needed was to be emphasized throughout.

Encouraged by growing support for such a project, the DOE initiated a genome program in fiscal year 1987 by redirecting funds from existing sources. The NIH followed suit in fiscal year 1988, when Congress appropriated earmarked funds to both agencies. Initial concern that dividing the project between two agencies could lead to unnecessary duplication and unproductive competition was allayed when the agencies decided that close cooperation and joint planning was of mutual benefit. The 1988 Memorandum of Understanding has led to a highly successful cooperative effort between the agencies. DOE has brought to the table extensive expertise in computer science and engineering, while NIH has strength in genetics, molecular biology, and medicine.

Management Structure

The National Center for Human Genome Research (NCHGR), a distinct administrative unit reporting to the Director of the National Institutes of Health, was established by the Secretary of the Department of Health and Human Services on October 1, 1989. The center is responsible for the planning, coordination, and funding of all genome project research at NIH. Overall scientific advice and evaluation of progress toward achieving the goals is provided by the Program Advisory Committee on the Human Genome. All applications assigned to the center are peer reviewed in the customary NIH manner. Regular research grants are reviewed by the Division of Research Grants, by the NIH, and by special initiatives by the Center's own review committee. The center also has its own National Advisory Council, which provides the second level of review as well as overall advice on the management of the center.

The Department of Energy's genome program is housed in the Office of Health and Environmental Research, within the Office of Energy Research. Project advice is provided by the Health and Environmental Research Advisory Committee, and coordination is enhanced by the Human Genome Coordinating Committee. Peer review is carried out by ad hoc groups constituted for that purpose.

In order to assure adequate coordination between the NIH and DOE programs, a joint subcommittee of the NIH and DOE advisory groups, the Joint DOE-NIH Subcommittee on the Human Genome, has been formed to deal with issues of mutual interest. In addition, a number of joint working groups have been established, which report to the joint subcommittee and assist the advisors in gathering detailed information on the state of the art in various areas and in evaluating specific scientific strategies. Currently, there are four working groups in addition to the joint subcommittee. These are the Mapping Working Group, the Sequencing Working Group, the Working Group on the Mouse, and the Ethical, Legal, and Social Implications (ELSI) Working Group. Most of the working groups have a core membership but recruit other experts as needed. Working groups are formed or disbanded according to the dictates of the evolving scientific opportunities and needs.

Scientific Strategy

A scientific plan for the U.S. Human Genome Project, with specific goals for the first 5 years, was developed jointly by NIH and DOE and was published in the spring of 1990 (U.S. Department of Health and Human Services and U. S. Department of Energy 1991). The goals were set on the assumption that funding for the program would cost approximately $200 million per year for the combined NIH and DOE genome efforts. A lesser level of support may cause completion of the goals to be delayed beyond 5 years. This plan now guides funding and research activities. Over 10,000 copies of the plan have been distributed to scientists, administrators, and other interested parties around the world. The official 5-year period covered by the plan was set to begin on October 1, 1990, and to end on September 30, 1995. It represents the first third of the 15-year time frame in which the project is to be completed.

Development of the 5-year plan required not only consensus building among the scientific community involved in the project, but close coordination between the NIH and DOE programs. The initial ideas were developed at a retreat of NIH and DOE advisors in August, 1988, and were refined at a number of subsequent meetings. Several broad scientific areas that were deemed critical to the accomplishment of the Human Genome Project were identified, and, for each area, an ambitious but realistic goal that should be reached in 5 years was selected.

The precise statement of the goals is given in the Appendix. Some of the highlights, with special emphasis on the NIH program, follow below. It is worth noting here, for the sake of clarity, that although the NIH and DOE programs are very similar, two of the goals are not shared but are primarily the province of NIH. These are the human genetic linkage map and the study of model organisms.

Linkage Maps of Human Chromosomes

Linkage maps are essential for locating disease genes on chromosomes. The goal for this area is to develop maps consisting of markers spaced no more than 5 cM apart with an average spacing of 2 cM. All markers will be converted to sequence tagged sites (STS) (Olson et al. 1989) to promote ease of use; to foster integration of the linkage, physical, and sequence maps; and so that the complete information for use of the markers can be stored electronically.

Initially, a framework map with highly polymorphic markers spaced at an average interval of 10 cM will be developed. Such a map will require approximately 300 high-quality markers. As of January 1992, almost 300 candidate framework markers had been isolated and mapped, about 40% of which are DNA-sequence based and can be assayed by PCR. These markers, however, are not as evenly distributed as needed. Some chromosome maps are well on the way to completion, while clusters and gaps remain on others. Thus new markers are still being isolated and evaluated for their ease of use and evenness of distribution. The desired maps should be completed within 1 or 2 years. A catalog of the framework-quality markers that have been identified to date is available from NCHGR.

Physical Maps of Human Chromosomes

Physical maps allow access to the DNA in a given region of the genome. The most useful physical maps for this purpose are overlapping sets of cloned DNA (contigs) whose positions on the chromosome are known. A major portion of the human genome should be available in mapped contigs at the end of the first 5 years. The contigs will be identified by STS markers spaced at intervals of approximately 100,000 bp, so that the mapping information can be stored in data bases and so that the cloned DNA does not need to be stored for prolonged periods of time. Currently, contigs that are at least 2 million bp long, some of them as long as 8 million bp, are being generated in a number of laboratories. This is a gratifying advance over just last year, when the first 2- million-bp contig was announced.

Sequencing

Current methods for DNA sequencing are too expensive and too time-consuming to allow us to contemplate large-scale sequencing of the human genome. Therefore, the initial goal for sequencing is to improve the technology so as to reduce the cost to less than \$0.50/bp. Current cost is estimated to be \$2–\$5/bp. Research on new technologies, as well as research on the scaling up and improving of existing technology, is being supported. In the course of attempts to scale up current technology, the genomes of some model organisms, as well as especially interesting regions of human DNA, will be sequenced.

Model Organisms

To understand human genetic information, it is important to conduct parallel studies on selected model organisms. Physical and genetic maps of several organisms—*Escherichia coli,* yeast, and *Caenorhabditis elegans*—are already essentially complete, and pilot sequencing has begun. *Drosophila* is not far behind, with physical mapping well under way and the start of sequencing under discussion. Even at current costs, the DNA sequence of an organism such as *E. coli* or yeast would be well worth knowing, because the information content is so much more dense than it is in human DNA. Furthermore, many of the genes in these model organisms have enough homology to the corresponding human genes to be useful as probes.

It is reasonable to expect that, at the end of the 5 years, *E. coli* will be completely sequenced and that the sequencing of yeast, *C. elegans,* and *Drosophila* will be well underway. Genetic and physical mapping of the mouse genome is also getting started in an organized manner. Thus, in the next 5 years, an extraordinary wealth of genetic information from model organisms should become available.

Informatics

The magnitude of the information being produced by the Human Genome Project poses a challenge to the data base and analytical tools that are available. Within the first 5 years, research and development to bring these tools up to the level needed must take place hand-in-hand with the biological research. There is a critical need for computer systems that can support large-scale mapping and sequencing efforts. In addition, there must be effective public data base structures

in which final information can be stored, organized, and made available to the diverse community of users, in a form that allows ready comparison between mapping, sequencing, and related biological information.

Several publicly available data bases are currently serving as the main repositories of genomic information. GenBank is the major repository for sequence data. It has been supported initially by the National Institute of General Medical Sciences (NIH) and will be funded in the future by the National Library of Medicine (NIH). The Genome Data Base (GDB) is the prime public repository of human mapping information. Funding for GDB was recently taken over by the NIH and DOE genome programs, and there are future prospects for international participation in its management and funding. Both data bases need to expand their capabilities so that they will become truly representative of the complexity of genomic information, and they need to be linked to each other in a transparent way. Work has begun on a mouse data base (G Base and the Encyclopedia of the Mouse). Yeast, *Drosophila,* and *E. coli* data bases are under discussion. Because data bases will be the ultimate repositories of genomic information, the Human Genome Project considers the development and support of appropriate genomic data bases an integral part of its mission.

Ethical, Legal, and Social Considerations

The availability of increasing numbers of genetic tests for diseases or other biological characteristics of humans raises a number of social, legal, and ethical issues. A novel aspect of the genome project is that it will study these issues alongside the scientific ones. First conceived by NIH, the concept of including an ELSI component in the Human Genome Project has been adopted by the DOE, as well as by the programs of several other countries. It is hoped that problems can be anticipated and that policy options regarding how to deal with them can be developed. Three principle areas have been identified for initial study: privacy of genetic information, protection from discrimination based on genetics, and safe introduction of genetic tests into mainstream medical practice. In furtherance of the third of these goals, NCHGR recently awarded grants to a group of investigators who will evaluate a variety of approaches to cystic fibrosis carrier testing. Goals in this area also include both improving public and professional education in genetics and broadly involving the public, through discussions of the issues.

Research Training

As in any new endeavor, the genome project needs to train scientists in the new skills that will be needed to complete the project and to make use of the information that is produced. A training program for predoctoral and postdoctoral students, as well as for experienced scientists, has been initiated. Particular emphasis is being given to interdisciplinary training that creates bridges between biological and other sciences, such as computer science, engineering, and mathematics. A variety of specialized courses are also being developed to make access to new technologies available to the broader community.

Technology Development

The creation of new and improved technology will always be an important aspect of the genome project. Many processes are being automated and made amenable to the use of much smaller quantities of materials. Multiplexing and other methods of processing many samples at once are being explored, with reduction of costs being a key concern throughout. Both new biological methodologies and new instrumentation, as well as new computer strategies, are being pursued.

While novel developments are encouraged, the genome program seeks to remain flexible and open to the possibility that some entirely new method that makes current technology obsolete may come along. In fact, the anticipation that this would happen was built into the original cost and time projections for the project, and, if the 15-year goals are to be reached within cost, new or significantly improved technology will be essential.

Funding

The funding available to NIH and DOE up to the present time is shown in table 1. Fiscal years 1988 and 1989 can be viewed as the start-up years, when the 5-year plan was created and when the various funding mechanisms needed for getting the project done were established. Since, by fiscal year 1991 (October 1, 1990), the funding level had risen sufficiently close to the estimated $200 million per year that was required to complete the project, that year was declared the first year of the official 15-year period.

Although the original cost estimates are now almost 4 years old, recently revised estimates still indicate that

Table 1

Funding for the Human Genome Project

Fiscal Year	HGP Budgets (millions of dollars) NIH	DOE[a]
1988	17.2	10.7
1989	28.2	18.5
1990	59.5	27.8
1991	87.4	47.7
1992 (estimated)	104.8	59.0

[a] Does not include salaries and expenses of DOE employees devoted to this effort.

essentially the same total level of funding is needed if the project is to be completed in 15 years. While individual cost items have changed, the overall level of effort involved is still the same. All these estimates assume, of course, that the cost of sequencing DNA will drop by at least a factor of 10.

International Collaboration

The Human Genome Project is not solely a U.S. initiative, although the United States was the first country to make a financial commitment to the project and, thus, currently has the largest and most advanced program. Increasing interest is being shown by other countries around the world. Substantial programs are currently underway in the United Kingdom, France, and the European Community, as well as in Japan. Smaller programs are in effect or expected in several other European countries and in Russia. Canada also has a program under discussion.

Coordination of these diverse efforts is clearly a challenge. Although NIH and DOE have kept in close contact with developments abroad and have established good working relationships with foreign agencies, such efforts will become increasingly time-consuming as the number of international programs grows. It is for these reasons that scientists organized a new society called the "Human Genome Organization," or "HUGO" for short. This organization is now functioning with regional offices in the United States, the United Kingdom, and Japan. International coordination and the setting of guidelines for sharing of materials and information are some of the areas of keen concern to HUGO.

In the future, HUGO expects to take on a greater role in the coordination of the human chromosome workshops, which were started by NIH and DOE over the past 2 years. These promise to become the major mechanism for the creation of consensus maps—both linkage and physical—and for the correlation of such maps with functional information. Meanwhile, the traditional Human Gene Mapping meetings will evolve to take account of the role both of the chromosome workshops and of the more sophisticated computer and data-base capabilities that are available today.

Conclusion

The Human Genome Project is enormously challenging from several points of view: scientific, technological, organizational, international, and sociological. It was initiated, after many years of debate, in response to strong recommendations by the scientific community. Although it was first considered by many to be too complex and difficult, the faith of the project's initial proponents has been shown to be justified. Progress on all fronts has been remarkable, confirming the optimism and judgment of early reviewers, such as the NRC and the OTA committees, that the time was ripe for a concerted effort to map and sequence the human genome. The project appears more accomplishable with every passing year. It has created a remarkable momentum in the scientific community and, indeed, internationally. Already, the methods and resources developed under this program are affecting and expediting many research projects. It represents the cutting edge of science, an investment in long-term technology development and cost reduction and an extraordinary promise for understanding human health and disease. With the tools and resources developed under the Human Genome Project, human genetics will come into its own and take on a central role in the study of human biology, a role that genetics has already played in the study of model organisms. If sufficient funding is appropriated, there is optimism that the challenging goals that the project has set for itself can be met.

Appendix

Five-Year Goals of the Human Genome Project

Mapping and Sequencing the Human Genome

Genetic map

1. Complete a fully connected human genetic map with markers spaced an average of 2–5 cM apart.
2. Identify each marker by a sequence-tagged site (STS).

Physical map

1. Assemble STS maps of all human chromosomes with the goal of having markers spaced at approximately 100,000-bp intervals.
2. Generate overlapping sets of cloned DNA or closely spaced unambiguously ordered markers with continuity over lengths of 2 million bp for large parts of the human genome.

DNA sequencing

1. Improve current methods and/or develop new methods for DNA sequencing that will allow large-scale sequencing of DNA at a cost of $0.50/bp.
2. Determine the sequence of an aggregate of 10 million bp of human DNA in large continuous stretches in the course of technology development and validation.

Model Organisms

Prepare a genetic map of the mouse genome, based on DNA markers. Start physical mapping on one or two chromosomes.

Sequence an aggregate of about 20 million bp of DNA from a variety of model organisms, focusing on stretches that are 1-million-bp long, in the course of the development and validation of new and/or improved DNA sequencing technology.

Informatics—Data Collection and Analysis

Develop effective software and data-base designs to support large-scale mapping and sequencing projects.

Create data-base tools that provide easy access to up-to-date physical mapping, genetic mapping, chromosome mapping, and sequencing information and that allow ready comparison of the data in these several data sets.

Develop algorithms and analytical tools that can be used in the interpretation of genomic information.

Ethical, Legal, and Social Considerations

Develop programs addressed at understanding the ethical, legal, and social implications of the project.

Identify and define the major issues and develop initial policy options to address them.

Research Training

Support research training of pre- and postdoctoral fellows starting in fiscal year 1990. Increase the numbers of trainees supported, until a steady state of about 600/year is reached by the fifth year.

Examine the need for other types of research training in the next year.

Technology Development

Support innovative and high-risk technological developments as well as improvements in current technology, to meet the needs of the project as a whole.

Technology Transfer

Enhance the already close working relationships with industry.

Encourage and facilitate the transfer of technologies and of medically important information to the medical community.

References

National Research Council Committee on Mapping and Sequencing the Human Genome (1988) Mapping and sequencing the human genome. National Academy Press, Washington, DC

Office of Technology Assessment (1988) Mapping our genes: genome projects: how big, how fast? Congress of the United States, Office of Technology Assessment, Washington, DC

Olson M, Hood L, Cantor C, Botstein D (1989) A common language for physical mapping of the human genome. Science 245:1434–1435

U.S. Department of Health and Human Services and U. S. Department of Energy (1990) Understanding our genetic inheritance, the U.S. Human Genome Project: the first five years FY 1991–1995. NTIS, U.S. Department of Commerce, Springfield, VA

Watson, JD (1990) The Human Genome Project: past, present, and future. Science 248:44–49